

6. FUTURE WORK

My database still needs some taxonomic consolidation. For example, I still have to assess the type-species of many of the genera as well as the names involved in many nomina nova.

An immediate target is to finish entering the classification of WOLTERS, but the ultimate goal of my taxonomic referential database is to contain all names of birds that have been published until today. Therefore the next logic step would be to computerize the Catalogue of Birds in the British Museum (SHARPE 1874-1898). Given the different quality of the taxonomic information varying from author to author of the 27 volumes of this enormous work, any added name needs careful checking of its reference. It is difficult to estimate the time outlay for this work.

A different and/or additional way to increase the completeness of my database is to continue entering type data of different museums world-wide. A first trial with the digitalized type data of the NHM was a great success: more than 80% of all names could be matched with taxa already entered into the database.

Obviously, further co-operation with additional museums possessing types is welcome. In return each museum would be allowed to access my database, which will be housed on the MNHN web site.

Acknowledgments

First of all, I want to thank Eric Pasquet for his continuous encouragement and his help introducing me into the world of ornithologists. Secondly, I thank Robert Prys-Jones and Frank Steinheimer for their co-operation in sharing the type data of the NHM, and Frank for his careful reading of the French/English draft of this paper and Edward C. Dickinson for his critics and advice. I also address my thanks to Renate van den Elzen for inviting me to present my talk at the Bonn conference.

Literature

CODD, E. F. (1970): A relational model of data for large shared databanks. *CACM* **13** (6).

DEVILLERS, P. & OUELLET, H. (1993): Noms français des oiseaux du monde avec les équivalents latins et anglais. Sainte-Foy, Québec, Canada: Éditions MultiMonde inc.; Bayonne, France: Éditions Chabaud. xiv+452 pp.

HOWARD, R. & MOORE, A. (1980): A complete checklist of the birds of the world, Oxford: Oxford University Press, viii + 701 pp.

HOWARD, R. & MOORE, A. (1984): A complete checklist of the birds of the world, Rev. ed., London: Macmillan, viii + 732 pp.

HOWARD, R. & MOORE, A. (1994): A complete checklist of the birds of the world, 2nd ed., London: Academic Press, xxxiv + 630 pp. ISBN 0-12-356910-9.

ICZN (1999): International Code of Zoological Nomenclature. 4th edition. International Trust for Zoological Nomenclature, London, XXIX + 306 pp. ISBN 0-85301-006-4.

MAYR, E. & COTTRELL, G. W., eds. (1979): Check-list of Birds of the World. Revision of the Work of James L. Peters, Vol. **I**, 2nd edition, XVII+547 pp. Cambridge, Massachusetts: Museum of Comparative Zoology.

MAYR, E. & COTTRELL, G. W., eds. (1986): Check-list of Birds of the World. A Continuation of the Work of James L. Peters, Vol. **XI**, XII+638 pp. Cambridge, Massachusetts: Museum of Comparative Zoology.

MAYR, E. & GREENWAY, J. C., jr., eds. (1960): Check-list of Birds of the World. A Continuation of the Work of James L. Peters. Vol. **IX**, XII+506 pp. Cambridge, Massachusetts: Museum of Comparative Zoology.

MAYR, E. & GREENWAY, J. C. jr., eds. (1962): Check-list of Birds of the World. A Continuation of the Work of James L. Peters. Vol. **XV**, X+315 pp. Cambridge, Massachusetts: Museum of Comparative Zoology.

MAYR, E. & PAYNTER, R. A. jr., eds. (1964): Check-list of Birds of the World. A Continuation of the Work of James L. Peters. Vol. **X**, IX+502 pp. Cambridge, Massachusetts: Museum of Comparative Zoology.

PAYNTER, R. A. jr., ed., in consultation with E. MAYR (1967): Check-list of Birds of the World. A Continuation of the Work of James L. Peters. Vol. **XII**, IX+495 pp. Cambridge, Massachusetts: Museum of Comparative Zoology.

PAYNTER, R. A. jr., ed., in consultation with E. MAYR (1968): Check-list of Birds of the World. A Continuation of the Work of James L. Peters. Vol. **XIV**, X+433 pp. Cambridge, Massachusetts: Museum of Comparative Zoology.

PAYNTER, R. A. jr., ed., in consultation with E. MAYR (1970): Check-list of Birds of the World. A Continuation of the Work of James L. Peters. Vol. **XIII**, XIV+443 pp. Cambridge, Massachusetts: Museum of Comparative Zoology.

PAYNTER, R. A. jr. (1987): Check-list of Birds of the World, Comprehensive index, XI+550 pp. Cambridge, Massachusetts: Museum of Comparative Zoology.

PETERS, J. L. (1931-1951): Check-list of Birds of the World. Cambridge, Massachusetts: Harvard University Press. (reprinted, Cambridge, Massachusetts: Museum of Comparative Zoology)

Vol. I, XVIII+345 pp., 1931, reprinted 1979 (cf. MAYR & COTTRELL 1979).

Vol. II, XVII+401 pp., 1934, reprinted 1963 ;

Vol. III, XIII+311 pp., 1937, reprinted 1961 ;

Vol. IV, XII+291 pp., 1940, reprinted 1964 ;

Vol. V, XI+306 pp., 1945, reprinted 1955 ;

Vol. VI, XI+259 pp., 1948, reprinted 1964; Cambridge, Massachusetts: Museum of Comparative Zoology

Vol. VII, X+318 pp., 1951.

Vol. VIII see TRAYLOR (1979)

Vol. IX see MAYR & GREENWAY (1960)

Vol. X see MAYR & PAYNTER (1964)

Vol. XI see MAYR & COTTRELL (1986)

Vol. XII see PAYNTER (1967)

Vol. XIII see PAYNTER (1970)

Vol. XIV see PAYNTER (1968)

Vol. XV see MAYR & GREENWAY (1962)

Vol. XVI see PAYNTER (1987)

SHARPE, R. B. (Ed., 1874-1898): Catalogue of Birds in the British Museum. 27 vols., London

Vol. I, by R. B. Sharpe, XIII+479 p., 14 pls. col., text ill., 1874.

Vol. II, by R. B. SHARPE, XI+325 p., 14 pls. col., text ill., 1875.

- Vol. III, by R. B. SHARPE, XII+343 p.: 14 pls. col., text ill., 1877.
- Vol. IV, by R. B. SHARPE, XVI+494 p., 14 pls. col., text ill., 1879.
- Vol. V, by H. SEEBOHM, XVI+426 p., 18 pls. col., text ill., 1881.
- Vol. VI, by R. B. SHARPE, XIII+420 p., 18 pls. col., text ill., 1881.
- Vol. VII, by R. B. SHARPE, XVI+698 p., 15 pls. col., text ill., 1883.
- Vol. VIII, by H. GADOW, XIII+385 p., 9 pls. col., text ill., 1883.
- Vol. IX, by H. GADOW, XII+310 p., 7 pls. col., text ill., 1884.
- Vol. X, by R. B. SHARPE, XIII+682 p., 12 pls. col., text ill., 1885.
- Vol. XI, by Ph. L. SCLATER, XVI+431 p., 18 pls. col., text ill., 1886.
- Vol. XII, by R. B. SHARPE, XV+871 p., 16 pls. col., text ill., 1888.
- Vol. XIII, by R. B. SHARPE, XVI+701 p., 15 pls. col.: text ill., 1890.
- Vol. XIV, by Ph. L. SCLATER, XIX+494 p., 26 pls. col., text ill., 1888.
- Vol. XV, by Ph. L. SCLATER, XVII+371 p., 20 pls. col., text ill., 1890.
- Vol. XVI, by E. HARTERT, XV+703 p., 14 pls. col., text ill., 1892.
- Vol. XVII, by R. B. SHARPE & W. R. OGILVIE-GRANT, XI+522 p., 17 pls. col., 1892.
- Vol. XVIII, by E. HARGITT, 597 p., 14 pls. col., text ill., 1890.
- Vol. XIX, by G. E. SHELLY, XII+484 p., 13 pls. col., text ill., 1891.
- Vol. XX, by T. SALVADORI, XXVI+658 p., 18 pls. col., text ill., 1891.
- Vol. XXI, by T. SALVADORI, XVII+676 p., 15 pls. col., 1893.
- Vol. XXII, by W. R. OGILVIE-GRANT, XVI+585 p., 8 pls. col., 1893.
- Vol. XXIII, by R. B. SHARPE, XIII+353 p., 9 pls. col., 1894.
- Vol. XXIV, by R. B. SHARPE, XII+794 p., 7 pls. col., text ill., 1896.
- Vol. XXV, by H. SAUNDERS, 1-339 and by O. SALVIN, 340-455; XV+475 p. 8 pls. col., 1896.
- Vol. XXVI, by R. B. SHARPE and W. R. OGILVIE-GRANT, XVII+687 p., 14 pls. col., 1898.
- Vol. XXVII, by T. SALVADORI, 636 p., 19 pls. col., 1895.
- SIBLEY, C. G. & B. L. MONROE, Jr. (1990): *Distribution and Taxonomy of Birds of the World*, Yale University Press, New Haven & London XXIV+1111 pp., ISBN 0-300-04969-2.
- SIBLEY, C. G. & B. L. MONROE, Jr. (1993): *A supplement to Distribution and Taxonomy of Birds of the World*, Yale University Press, New Haven & London, VI+108 pp., ISBN 0-300-05549-8.
- TRAYLOR, M. A. jr., ed. (1979): *Check-list of Birds of the World, A Continuation of the Work of James L. Peters*. Vol. VIII, XV+365 pp. Cambridge, Massachusetts: Museum of Comparative Zoology.
- WOLTERS, H. E. (1975-1982): *Die Vogelarten der Erde: eine systematische Liste mit Verbreitungsangaben sowie deutschen und englischen Namen*, Hamburg: Verlag Paul Parey XX+748 p., ISBN 3490091183.

Rémy BRUCKERT, Muséum National d'Histoire Naturelle, Département systématique et évolution - Unité taxonomie et collections, 55, rue Buffon, 75005 Paris, France, e-mail: bruckert@mnhn.fr

Computerizing Bird Collections and Sharing Collection Data Openly: Why Bother?

A. Townsend PETERSON & Adolfo G. NAVARRO-SIGÜENZA

Natural History Museum, Lawrence, Kansas, USA and Museo de Zoología, México, México

Abstract. Natural history museum collections provide the basic documentation of life on Earth. As such, they represent the critical and unique resource by which that life may be understood, and have immense economic and scientific importance. Nevertheless, particularly in recent decades, natural history museums have seen less and less attention – and resources – in spite of their importance. A series of new efforts, however, aim to recoup that prominence via community efforts to unite data resources towards a vastly improved understanding of biodiversity and its implications. The Species Analyst represents an effort to unite natural history collections databases worldwide to this end: 77 institutions now cooperate or are committed to cooperate in serving records of 51 million natural history museum specimens to users worldwide, and has seen more than 700,000 users to date.

Key words. The Species Analyst, TSA, archive for biodiversity, worldwide facilities

1. INTRODUCTION

Computerization of ornithological collections is increasingly considered a priority for curators and staff of natural history museums. A common quandary, however, is how and why to get started. The curator is presented with a bewildering variety of databasing programs, some especially designed for specimen records, and others off-the-shelf generic database programs that can be customized for any use. Choice of a platform, choice of data fields, and choice of computerization strategy all become critical – and difficult – consideration. Unfortunately, these considerations can often seem so complex that computerization efforts are not initiated.

Moreover, presented with a thousand and one other priorities of collections building, specimen conservation, institutional politics, and research efforts, and given the significant time investment that computerization requires, the question arises as to whether the result is worth the time. That is, one must consider what are the benefits of computerization, and how much do they benefit the collection, the curator, and the broader community.

The purpose of this contribution is to provide a rationale for computerizing bird collections as a critical step forward in their care. Along the way, we review steps involved – a sort of minimum-standard guide to starting computerization efforts. Finally, we provide a series of examples of how computerizing collections data, and sharing those data across many institutions worldwide, benefits the collections themselves.

2. WHY COMPUTERIZE A COLLECTION?

Databasing or computerizing a collection is a lot of work, and may easily absorb years of effort. So why do it? Several reasons argue strongly for taking this step. A partial list follows:

- *Get to know your collection* – a sweep through the whole collection, drawer by drawer, gives a unique knowledge of a particular collection.
- *Discover important specimens* – many fascinating discoveries have resulted from the specimen-by-specimen attention during computerization efforts, including species new to science, lost type specimens, important historical specimens, etc.
- *Detect problems* – again, the specimen-by-specimen attention can help to detect serious problems that might otherwise not be noticed ... damage from insects or water, fading of plumages, drying of spirit specimens, etc.
- *New views of the collection* – although we are familiar with summaries of collections in terms of taxonomic completeness, and perhaps regional summaries, many new views of collections open when a collection is computerized, e.g., maps of the geographic distribution of specimens, summaries of accessions over time, etc.
- *Save curatorial time* – making summaries of holdings, preparing loan invoices, tracking down particular specimens, and many other curatorial tasks are considerably more efficient when the collection is available in database form.
- *Standardize taxonomy* – once data are in electronic form, comparing names against a standard list (e.g., the Peters' check-list) can identify a first set of non-

- standard names that require checking and updating.
- *Efficient information access* – many questions and data requests that require hours or days of work for an uncomputerized collection will suddenly become feasible to answer in minutes, making possible much more creative uses of the information in collections. For example,
 - What are your holdings of taxon X?
 - What are your holdings from country X?
 - Do you have specimens collected by person X?
 - What is the history of specimen acquisition rates in your collection?
 - And many more ...

In short, computerization of a collection is a major undertaking, but ends up repaying the investment of time and effort many times over.

3. CHOOSING A PLATFORM

The first big question to be answered is about which platform (databasing program) to use. This decision becomes complex ... sometimes, museum administrators decide to force all collections in the museum to use the same program. Even if one has the freedom to choose, should one choose among the many programs that have been developed specifically for natural history museum specimens (BIOTA, BIOTICA, SPECIFY, etc.), or a generic program off the shelf (e.g., Microsoft Access, Ora-

cle)? Regarding this choice, each option has its strengths and weaknesses (Table 1). In general, we would recommend the off-the-shelf option for small, old, or inactive collections, and the specimen databasing programs for larger, data-rich, and very active collections.

Regardless of this choice, one should insist on several minimum criteria for a databasing platform. These criteria are critical features of a program that must be fulfilled in order to avoid problems. As follows:

- Capacity for export to other, generic formats, particularly ASCII delimited format, to allow reporting, export to other programs, and porting to future technologies and platforms.
- Compatible with Standardized Query Language (SQL), which permits many functionalities to be added to your database related to sharing data.

Once a platform has been identified that fits the particular needs of a collection, and meets these basic requirements, then design of the computerization effort can begin.

If the reasoning outlined above suggests that the best solution to computerization is that of a more complex program specifically designed for natural history specimen data, then you should read about several of the programs that are available. Links to a number of such programs are presented in Table 2.

Table 1: Summary of advantages and disadvantages of specialized versus generic programs as platforms for computerizing bird collections.

NATURAL HISTORY MUSEUM SPECIMEN DATABASING PROGRAMS	OFF-THE-SHELF GENERIC DATABASING PROGRAMS
<p>Advantages</p> <p>Designed specifically for specimen management</p> <p>Features such as authority lists, loan invoice reporting, etc.</p> <p>No customization or little customization required</p> <p>Most complex solutions specific to natural history specimens are tractable</p>	<p>Long-term continuity of support from the company</p> <p>Easy availability of expert advice, given broad usage in many communities</p> <p>Simplest solutions are feasible</p> <p>Simple learning curve</p>
<p>Disadvantages</p> <p>Can disappear – long-term support often depends on a person – researcher or developer – who can decide not to support the program further, or who may decide not to update to newer versions (e.g., MUSE)</p> <p>Expert advice may be unavailable in a particular city</p> <p>May not permit very simple solutions to simple problems</p> <p>Steeper learning curve</p>	<p>May need customization of program for intermediate-to-complex situations</p> <p>Not designed specifically for specimen data</p> <p>Complex features (e.g., reporting, authority lists) not automatically available</p>

Table 2: Selected specialized programs designed specifically for collections data. Provided are World Wide Web links for more information.

Program	URL
SPECIFY	http://usobi.org/specify/
Biótica	http://www.conabio.gob.mx/biotica_ingles/distribucion_b.html
BioLink	http://www.biolink.csiro.au/
BIOTA	http://viceroy.eeb.uconn.edu/biota
KE EMu (not recommended for integration via <i>Species Analyst</i>)	http://www.kesoftware.com/

4. CHOOSING DATA FIELDS

This step may prove to be the most critical of all in the process of computerization. With too many fields, time and filespace are wasted, whereas with too few, they will have to be added later or one will have to live without them. If an incorrect structure is chosen, the database may be forever handicapped by this design flaw. However, the challenge is reduced quite a bit with an understanding of a few basic ideas. Specimen data, in their simplest form, distill down to three linked sets of information about each specimen:

- *Taxonomic information* – the taxonomic identity of the specimen
- *Geographic information* – the geographic location of its collection
- *Detailed documentation of the specimen* – time of collection, collector identity, museum catalogue number, sex, age, body mass, etc.

Thinking in this manner, we can envision a structure for a specimen database that would capture this information optimally. Taxonomy and geography are both hierarchical concepts, and so we can represent them as such, which would make for three interacting sets of information (Fig. 1).

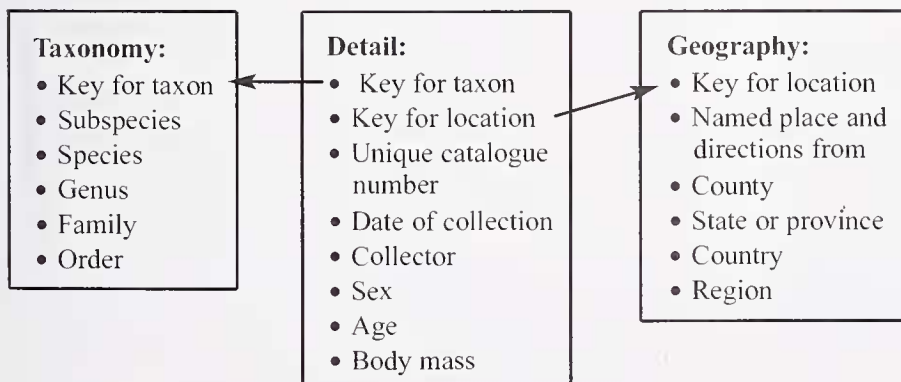


Fig. 1: Diagrammatic illustration of a simple relational database structure designed to link hierarchically organized geographic and taxonomic information with specific data regarding a particular specimen.

In the simplest sense, then, even in a spreadsheet program such as Microsoft Excel, or (better still) as a single table in a database program such as Microsoft Access, one could use a straightforward single table that holds critical fields (see Table 3). This very simple structure provides a clear, workable solution for small collections. In a more complex situation, in which more specimens are to be computerized, this structure can be made relational (Fig. 1) (that is, made up of several tables that interconnect). The advantage of a relational database structure is that elements of the database are entered only once: e.g., the locality descriptor for the 150 specimens collected at USA/Kansas/Douglas Co./Lawrence/10 km E is entered only once, reducing the possibility of typographical errors.

Table 3: Critical minimum set of fields for a simple collections database.

Field	Example
Catalogue number	15230
Genus	Cyanocitta
Species	cristata
Subspecies	cristata
Date of collection	24 October 1956
Collector	Fredrick E. Jones
Sex	Female
Age	Adult
Body mass	120 g
Country	USA
State or province	Ohio
County	Butler Co.
Named place and directions from	Oxford, 10 km E

This sort of simple relational structure can be implemented in a program such as Microsoft Access with a few hours' attention by a technician familiar with the program. The custom specimen database programs use a more complex relational structure, but one that is in essence based on this overall backbone. Again, the more complex the demands that one will wish to place on the database (e.g., more complex queries, more detailed reporting, more specimens), the more complex the database structure that will be required. For relatively simple applications, however, the simple flat file (single table) setup described above will often be adequate.

5. COMPUTERIZATION STRATEGY

The next question to be faced is the strategy for computerization. This decision depends heavily on the exact situation of a collection. If, on the one hand, an excellent paper catalogue or card file exists, one may wish to computerize directly from that, and then verify the accuracy and completeness later from the actual specimens. If, on the other hand, a good card file or catalogue does not exist, or if many specimens may have been omitted (exchanged or deaccessioned) or not entered in the catalogue, then you may be better off computerizing directly from specimens.

In general, two passes through the collection will be necessary as part of any computerization effort. The first will simply get each specimen's data into the computer as efficiently as possible. The second will verify (1) the existence of the specimen, (2) that all data elements are entered in the database, and (3) that all of the specimen's data are correct as entered. This verification step, although labor-intensive, is critical to making the database a correct representation of the information contained in the specimens' labels.

All computerization efforts should involve the critical step of backing up data at regular intervals. Too many 'impossible accidents' have removed a year of work, and set a computerization effort back terribly. Backing up data should be done as permanently as possible ... that is, compact disks are better than floppy disks. It should also be done with redundancy: each time that you make a copy, if at all possible, it should not over-write the previous copy. This preservation of 'versions' of the database allows one to go back a week or a month if some error appears in the data set. Finally, given the possibility of more catastrophic losses, the back-up copies should be stored off-site, preferably in several places. Excellent storage sites for these copies can include libraries or archives, or curators' homes, or they can even be transferred via the Internet or via mail to another country.

6. THE SPECIES ANALYST (TSA)

The Species Analyst (<http://speciesanalyst.net/>) is a collection of software tools that permits integration of computerized collections data among institutions around the world into a distributed biodiversity information facility. For example, a user might wish to ask for records of any taxon from Yellowstone National Park or from Burma, or all specimens collected by Alexander von Humboldt, and retrieve information in a matter of seconds from 50 institutions around the world.

TSA uses a hybrid of Z39.50 (an information transfer protocol developed about 20 years ago in the bibliographic community) and XML (a more modern and efficient protocol) to permit efficient query and

retrieval of data. TSA may be accessed via a web portal that permits basic queries, or via extensions to Microsoft Excel (for retrieval of data in spreadsheet format) and ArcView (for retrieval of data as GIS coverages) (downloads available at <http://speciesanalyst.net/downloads>).

TSA currently integrates data sets from 22 institutions, for a total of 15 million specimen data records for over 50,000 species; a total of 58 institutions has committed to participation formally, which will take the total number of specimen records served to about 50 million. A special strength at present is in ichthyological data, as FishNet (<http://speciesanalyst.net/fishnet/>) has taken excellent advantage of TSA technology to create a data facility linking most important computerized fish collections. Now funded is a parallel network for mammal collections data (MANIS, based at the Museum of Vertebrate Zoology; <http://elib.cs.berkeley.edu/manis/>), and networks for herpetological and ornithological (expanded) specimen data are pending and in preparation, respectively.

7. WHY SHARE DATA ONCE COMPUTERIZED?

Above, we listed the first set of benefits of computerization of bird collections – namely, freer and more complete access to the information content of the specimens that make up the collection. These benefits are indeed considerable, and add enormously to a curator's ability to take care of a collection. However, once data are computerized, if they are shared, and integrated with data from other collections around the world, an additional set of benefits accrues.

In essence, a set of emergent properties comes into being once all (or nearly all) data are integrated for a particular taxon or region. We have come to appreciate these emergent properties as we have assembled the Atlas of Mexican Bird Distributions (NAVARRO & PETERSON, in prep.), a centralized database now including the contents of more than 60 natural history museum collections of Mexican birds. This 11-year project has resulted in a diversity of synthetic publications regarding the Mexican avifauna (NAVARRO-SIGUENZA et al. 1992a, b; PETERSON 1993; PETERSON et al. 1993; PETERSON 1998; PETERSON et al. 1998a, b; NAVARRO-SIGUENZA & PETERSON 1999, 2000; PETERSON et al. 2000, 2001, 2002). Herein, we will use this exemplar data set to demonstrate a variety of potential benefits to broad integration of data across institutions, as follows:

7.1 Georeferencing as a Community

Georeferencing locality data for specimens opens doors to a multitude of new capabilities and new func-

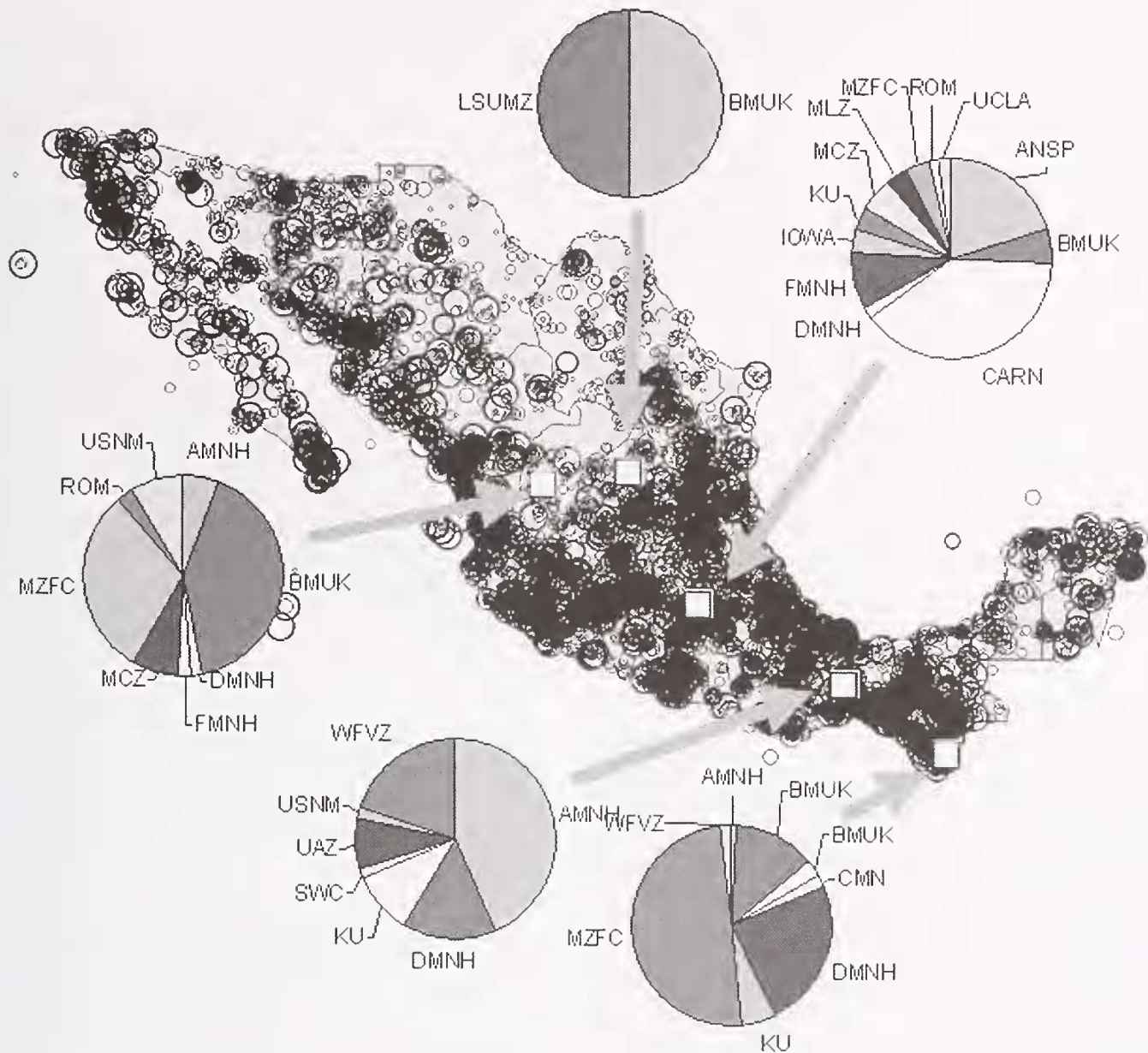


Fig. 2: Map of Mexico with collecting localities plotted by numbers of specimens collected at each point (graded symbol size: smallest = 1 specimen, largest = >100 specimens). For five points, to illustrate the redundancy of collecting localities among museums, we provide pie diagrams that illustrate the relative holdings of specimens from that particular site among scientific collections (see Acknowledgements for institutions and abbreviations).

tionalities to collections data. Indeed, all of the advances of geographic information systems (GIS) open up to collections data once latitude and longitude data are available for the collecting localities for each specimen. Nevertheless, georeferencing collections data – even once they are in electronic form – represents an enormous task.

Integrating this task over many institutions, however, takes advantage not just of having more people to help in a large task, but also of the redundant nature of the geographic sampling of birds (Fig. 2). Indeed, more than 25% of Mexican bird collecting localities occur in more than one museum, and some in more than 20 museums. This redundancy results from col-

lections being dispersed among numerous museums (e.g., the specimens of Wilmot W. BROWN from Chilpancingo, Guerrero), and from certain sites being especially accessible or well-known as collecting localities in particular regions (e.g., Cerro San Felipe, Oaxaca).

A first experiment in cooperative georeferencing is beginning in the mammal community in the United States. The MANIS network, a U.S. National Science Foundation-funded effort, is connecting 17 institutions with computerized holdings of mammal specimens. A first step in MANIS' integration efforts is the pooling of institutional lists of localities to be georeferenced; institutions are then 'signing up' for particu-

lar regions, perhaps a home state, or an area of particular interest to the curator. In this way, efforts in georeferencing have a direct return for a particular investigator or institution, and add to the community pool of georeferenced information.

7.2 Detecting Errors in Date and Locality

Once specimen data are integrated, and have been georeferenced, further data refinements are possible. A common question is that of the relative reliability of the data associated with specimens from different collectors (BINFORD 1989). Because of the fragmented and dispersed nature of collector's material it has always been out of reach before. For instance, the still-living collector and ornithologist Robert W. DICKERMAN has deposited specimens at 14 of the 32 museums included in our present summary; the early twentieth century collector Wilmot W. BROWN has specimens distributed across 23 of the 32 museums. Once these data are pooled, however, new insights become possible regarding collectors' relative reliability.

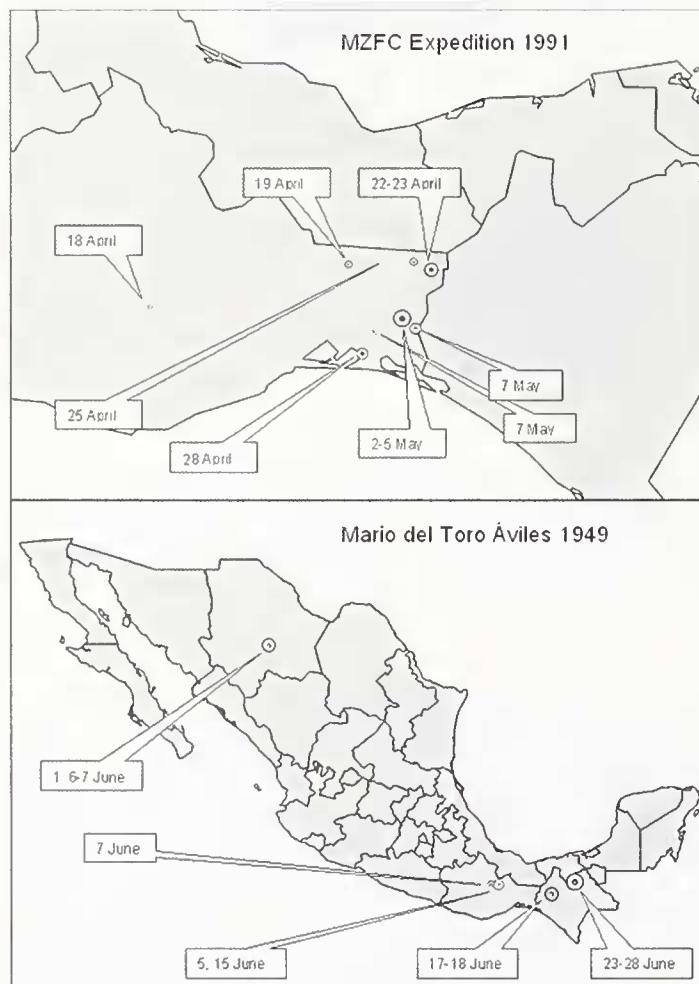


Fig. 3: Maps of collecting localities for two contrasting groups of collectors in Mexico: a Museo de Zoología, UNAM (MZFC) expedition in Spring 1991, and the collections of Mario del Toro Áviles in June 1949. Organized by collections date, consistencies and inconsistencies of specimen labeling become clear.

Basically, by assembling the entire opus of a collector, and sorting specimen locality by collecting date, it is possible to assess how geographically reasonable the combination of dates and localities is. Hence, to present a contrasting pair of examples, a Museo de Zoología, UNAM, expedition in 1991 scouted numerous sites in central and eastern Oaxaca (Fig. 3, top); although its route was complex, specimens from particular localities were clumped in time, and a sensible route could be reconstructed (although, in constructing this example, we detected an error in our georeferencing ... the 'Benito Juárez' referred to in the locality descriptor was the one in eastern Oaxaca, not the one in central Oaxaca). In stark contrast, specimens scattered across four museums (MLZ, LACM, FMNH, USNM) suggest that the infamous collector Mario del TORO ÁVILES worked at several sites across Mexico in June 1949; plotting these localities by date, however, reveals a number of points at which impossibly long journeys would have had to have been made in too short a time (Fig. 3, bottom). This result confirms earlier suspicions that del TORO ÁVILES' dates and localities are to be regarded with utmost caution (BINFORD 1989; PETERSON & NIETO-MONTES DE OCA 1996).

This approach can be used to detect problems in collectors' series, which will either be errors in date of collection or in collecting locality. Indeed, for an integrated, distributed data set consisting of the holdings of many institutions, it could be implemented as an error-seeking module that scans the data set collector by collector, and flags particular records as potential problems. These flagged specimen lists could then be distributed to collection curators for checking.

7.3 Detecting Errors in Identification or Georeferencing

A further refinement to specimen data also becomes possible, which will detect problems either in species identification or in georeferencing of localities. In essence, by viewing large quantities of occurrence data for a particular species, it is possible to detect spatial outliers, which likely represent identification or georeferencing problems. This process can be refined still further via ecological niche modeling for species: the ecological needs of a species are modeled (PETERSON 2001; PETERSON et al., in press) using high-end computational tools (STOCKWELL & NOBLE 1992; STOCKWELL 1999; STOCKWELL & PETERS 1999). These procedures use known occurrences of a species to produce a geographic view of areas meeting and not meeting its ecological needs; overlaying the same known occurrence points used to build the models allows identification of outlier occurrences.

As an example of this approach, we used the known occurrences of the brush-finch *Atlapetes pileatus* to